

Maciej WERYHO

Katarzyna MROTEK

Digitalizacja zbiorów i udostępnianie publikacji cyfrowych w Bibliotece Główniej Uniwersytetu Kazimierza Wielkiego w Bydgoszczy

Wstęp: społeczeństwo cyfrowe

Od kilkudziesięciu już lat technologie komputerowe i telekomunikacyjne w coraz większym stopniu generują zmiany w rozwoju naszej cywilizacji. Komputer osobisty czy telefon komórkowy nieodwracalnie zmieniają system komunikacji społecznej oraz stają się źródłem dostępu do ogromnych zasobów informacyjnych, które rosną z każdym dniem w sposób bezprecedensowy. Sieć globalna, zdecentralizowany zbiór różnorodnych obiektów cyfrowych, binarnych zapisów kultury i wszelkich przejawów ludzkiej aktywności, z jednej strony rodzi nadzieję na wolny i demokratyczny dostęp do wiedzy i edukacji, ale z drugiej niesie zupełnie nowe problemy i obawy. Powstanie i rozwój Internetu jest niewątpliwie jedną z ważniejszych rewolucji w ludzkich dziejach, porównywalną dziś do wynalezienia ruchomej czcionki i prasy drukarskiej w XV wieku. I tak, jak dzieło Jana Gutenberga rozpoczęło epokę powszechnego dostępu do wiedzy, umożliwiło dynamiczny rozwój edukacji, oświaty, kultury, i w ten sposób wpłynęło na wszystkie obszary życia kolejnych pokoleń, tak dzisiaj Internet staje się podstawowym medium niezwykle silnie wpływającym na nasze życie. Sieć bowiem jest nie tylko nowym i bardzo atrakcyjnym narzędziem pozyskiwania informacji, ale nadaje jej również nową formę, generuje nowe gatunki oraz energicznie stymuluje jej rozwój. Internet staje się obecnie jednym z podstawowych przejawów współczesnej kultury: sposobem komunikacji, ekspresji, rozrywki, pracy, zdobywania wykształcenia, wiedzy, inicjowania nowych ruchów społecznych, czy zawierania znajomości. Zmiany więc, jakie inicjują współczesne technologie sieciowe, wpływają na wszystkie dziedziny życia społecznego¹.

Niemal równocześnie z pojawieniem się komputerów osobistych połączonych w Sieć pojawiła się nadzieja na upowszechnianie drogą cyfrową dorobku ludzkiej myśli, zapisanej i przechowywanej w bibliotekach świata. Internet, od czasu gdy Tim Berners-Lee wyposażył

¹ P. Levinson, *Miękkie ostrze, czyli historia i przyszłość rewolucji informacyjnej*, Warszawa 2006, s. 33.

go w hipertekstowe łączy powołując do życia World Wide Web, a Marc Andreessen napisał Mosaic, pierwszą przeglądarkę pracującą w trybie graficznym, stał się masowo wykorzystywanym nośnikiem multimedialnym. Jednak mimo tak niezwyklej ewolucji form ekspresji w cyberprzestrzeni, które zmieniają się i przenikają wzajemnie, nadal podstawową pozostaje tekst. Informacje przechowywane dotychczas w postaci analogowej są więc konwertowane na postać binarną i transponowane do pamięci komputerów. W ten sposób zostają ponownie włączone w obieg komunikacyjny, tym razem cyfrowy. Treść zapisana w postaci numerycznej po raz pierwszy zostaje uwolniona i uniezależniona od dotychczasowego nośnika, co radykalnie zmienia możliwości jej udostępniania w Sieci². Każdy użytkownik Web otrzymuje do niej swobodny dostęp, niezależnie od jej rzeczywistej lokalizacji, czy ilości czytelników. Rodzi się więc utopiina wizja Internetu jako globalnej biblioteki cyfrowej, obejmującej zbiór piśmiennictwa światowego, kanon kultury i nauki, wirtualny odpowiednik Biblioteki Aleksandryjskiej, na dodatek w każdej chwili dla wszystkich otwartej.

Biblioteki cyfrowe

Już w 1971 roku Amerykanin Michael S. Hart wprowadził do pamięci komputera tekst pierwszej publikacji, przepisując, niczym średniowieczny kopista, Deklarację Niepodległości Stanów Zjednoczonych. Stała się ona początkiem Projectu Gutenberg, pierwszej kolekcji cyfrowej, która dzisiaj udostępnia użytkownikom Internetu ponad 20 000 tytułów³. Większość z nich wprowadzają do Sieci wolontariusze, którzy zrzeszeni w Literary Archive Foundation jednoczą się wokół wspólnej idei. Specyfiką tego pionierskiego projektu jest jego pełna niezależność od platformy systemowej, dlatego wszystkie publikacje prezentowane są w postaci czystych plików tekstowych. Każdy użytkownik sam może im nadać dowolną formę, stając się każdorazowo edytorem danej publikacji.

Takich kolekcji pełnotekstowych, oferujących dostęp nie tylko do pełnej zawartość książek, czasopism, czy artykułów, ale także grafik, zdjęć, wykresów, plików dźwiękowych, animacji, filmów, i wszelkich zbiorów przechowywanych w tradycyjnych bibliotekach, powstaje w Internecie coraz więcej. Przede wszystkim budowane są właśnie przez biblioteki.

² H. Holender, *Od dorywczej dygitalizacji do cyfrowego bibliotekarstwa*, w: *Dygitalizacja zbiorów bibliotecznych : materiały z ogólnopolskiej konferencji pt. "Dygitalizacja zbiorów bibliotecznych"*, Warszawa 3-4 czerwca 2005 r., red. Elżbieta Stefańczyk, Warszawa 2006, s. 7.

³ Project Gutenberg, dostępny w World Wide Web: [//www.gutenberg.org/wiki/Main_Page](http://www.gutenberg.org/wiki/Main_Page)

Bibliotekarze bowiem bardzo szybko otworzyli się na cyfrowe medium, i dostrzegli w nim nowe możliwości realizacji swojego podstawowego powołania⁴.

Największym narodowym projektem biblioteki cyfrowej jest American Memory, realizowany przez Bibliotekę Kongresu w Waszyngtonie, która udostępnia swoje zbiory online już od 1994 roku. W swojej kolekcji posiada cyfrowe odpowiedniki dokumentów o znaczeniu historycznym i kulturowym dla Ameryki, między innymi: fotografie z wojny secesyjnej, nagrania przemówień przywódców amerykańskich z okresu I wojny światowej, kolorowe fotografie z okresu II wojny światowej, a także broszury, rękopisy, nuty i kartografię. Użytkownik Sieci znajdzie również w zasobach American Memory wielotysięczny zbiór fotografii z archiwum „Chicago Daily News” z lat 1902-1933, reprodukcje ponad 900 plakatów, obszerną dokumentację ruchu emancypacyjnego w Stanach Zjednoczonych z lat 1848-1921, a nawet kilka zbiorów wczesnych, niemych filmów⁵.

Równie ciekawym i bardzo bogatym zbiorem cyfrowych dokumentów może pochwalić się działająca od 1999 roku Gallica la bibliothèque numérique. Jej zasoby przeznaczone są przede wszystkim dla francuskojęzycznej części świata, ale udostępnia także klasyczne dzieła w pierwotnych wersjach językowych. Obejmuje swoim zasięgiem multimedialne dokumenty od średniowiecza do początku XX wieku. Umożliwia dostęp do ponad 90 000 dokumentów cyfrowych, 80 000 obrazów oraz wielu plików dźwiękowych. Wszystkie jej zasoby są częścią Biblioteki Narodowej Francji. Zostały one wyselekcjonowane w taki sposób, aby zobrazować ogrom francuskiego dziedzictwa narodowego, a zarazem przedstawić wiedzę encyklopedyczną. Kolekcja dotyczy wielu dziedzin nauki, takich jak: historia, literatura, nauki przyrodnicze, filozofia, prawo, ekonomia, nauki polityczne. Gallica ma w swoim posiadaniu powieści, eseje, czasopisma, zbiory ikonograficzne, mapy oraz iluminowane rękopisy, co czyni ją obecnie jedną z bardziej znaczących bibliotek cyfrowych dostępnych w Internecie⁶.

Zasoby większości kolekcji udostępnianych online są jednak znacznie skromniejsze, co wynika z ograniczeń finansowych instytucji koordynujących dany projekt. Najczęściej tworzone są przez biblioteki narodowe i uniwersyteckie, jak chociażby Wright American Fiction 1851-1875, która udostępnia kolekcję tekstów literackich wyselekcjonowanych

⁴ D. Pearson, *Digitalizacja - czy istnieje jakaś strategia?*, „Biuletyn EBIB” 2003, nr 2, [on-line], [dostęp: 07.03.2007], dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2003/42/pearson.php>

⁵ American Memory, dostępny w World Wide Web: <http://memory.loc.gov/ammem/index.html>

⁶ Gallica la bibliothèque numérique, dostępny w World Wide Web: <http://gallica.bnf.fr/>

według drugiego tomu bibliografii Lyle H. Wright'a⁷, czy Perseus Digital Library, gromadząca kanon literatury humanistycznej⁸. Oczywiście nieporównywalnie większymi funduszami dysponują firmy komercyjne, które również podejmują już szerokie inicjatywy digitalizacji zbiorów, bądź to samodzielnie, bądź współfinansując dany projekt.

Dostęp zamknięty

Wydawało się więc, że jedynym problemem w powszechnym dostępie do wiedzy gromadzonej na dyskach komputerów będzie długotrwały i kosztowny proces konwersji zbiorów analogowych na postać binarną. Szybko jednak okazało się, że największa bariera ma zupełnie inny charakter. Stanowią ją mianowicie niezwykle surowe przepisy prawa autorskiego, które owszem, bardzo dobrze spełniają swoją funkcję w tradycyjnym rynku wydawniczym, natomiast zupełnie zmienia się obszar ich oddziaływań w obiegu cyfrowym. Sieć bowiem umożliwia swobodną wymianę informacji pomiędzy jej użytkownikami czy instytucjami, która z założenia nie ma i nie musi mieć charakteru komercyjnego. Niestety, obowiązujące obecnie regulacje prawne obejmują wszelkie działania generujące obieg informacji cyfrowej, nie różnicując ich, i w ten sposób drastycznie je ograniczają. W Polsce prawo autorskie chroni każdą publikację przez 70 lat od daty śmierci autora, a prawo wydawnicze przez 30 lat od daty jej publikacji. Materiały z tego obszaru nie mogą więc być w proces digitalizacji włączone. Konieczne jest każdorazowo negocjowanie warunków udostępnienia danego tytułu z autorem, wydawcą i podpisywanie w tym celu odpowiednich umów, co w praktyce przynosi znikome efekty. W konsekwencji zasoby bibliotek cyfrowych praktycznie w ogóle nie obejmują publikacji XX wiecznych. Sytuacja ta rodzi duże napięcie i dysonans pomiędzy możliwościami, jakie przyniosła globalna Sieć komputerowa, a ich rzeczywistą realizacją.

Najbardziej spektakularnym tego przykładem jest projekt stworzenia cyfrowego zasobu przez największego potentata doby Internetu, firmę Google Inc. Jest to dotychczas największa komercyjna inicjatywa digitalizacji zbiorów bibliotecznych. Przedstawiciele Google podpisali umowę z czterema amerykańskimi bibliotekami uniwersyteckimi: Stanforda, Harvardu, Oksfordu i Michigan oraz z Nowojorską Biblioteką Publiczną na skanowanie części ich księgozbiorów i stworzenie cyfrowej kolekcji w Internecie. Do końca

⁷ Wright American Fiction 1851-1875, dostępny w World Wide Web: <http://www.lettrs.indiana.edu/web/w/wright2/>

⁸ Perseus Digital Library, dostępny w World Wide Web: <http://www.perseus.tufts.edu/>

2015 r. planowane jest zeskanowanie 15 mln woluminów. Jednak wśród materiałów wyselekcjonowanych do projektu znalazło się także wiele pozycji współczesnych. Inicjatywa ta spotkała się więc od razu ze zdecydowanym protestem autorów oraz wydawców, którzy pozwali firmę Google do sądu. W efekcie, pod koniec 2005 r. serwis Google Book Search w wersji beta rozpoczął swoją działalność, udostępniając zeskanowane już materiały, jednak znacznie różni się on od pierwotnych założeń⁹. W szerokim zakresie udostępnione zostały jedynie publikacje wydane przed 1923 r., natomiast pozycje współczesne otwarte są wyłącznie dla mechanizmów wyszukiwawczych serwisu. Jego użytkownik ma jedynie możliwość pełnotekstowego przeszukiwania całego zasobu. A problemy prawne firmy cały czas trwają, im więcej tekstów pojawia się w Sieci, tym więcej spraw w sądzie. Ostatnio Google przegrała kolejny proces, tym razem z belgijskimi wydawcami prasy, którzy domagali się usunięcia z serwisu tekstów pochodzących z francuskojęzycznych gazet.

Na przykładzie działań Google'a widać więc bardzo wyraźnie, że obowiązujące obecnie prawo stanowi ogromną barierę w dostępie do informacji cyfrowych. Kolejny duży projekt digitalizacji zbiorów, realizowany przez konsorcjum Open Content Alliance, współtworzone m.in. przez Yahoo!, Adobe Systems oraz Microsoft, obejmuje już wyłącznie publikacje z obszaru public domain, czyli te, do których wszelkie prawa już wygasły¹⁰.

Dostęp otwarty

Swobodna wymiana informacji, wiedzy i wszelkich owoców kultury od początku istnienia Internetu stanowi jego integralną cechę i nadaje mu niepowtarzalny charakter. Bez wsparcia i pracy tysięcy osób udostępniających oprogramowanie z otwartym kodem źródłowym, wspólnie modyfikowane i nieustannie usprawniane, Web nigdy nie powstałby w takiej formie i w takim tempie. Otwartość stanowi jego siłę, to bowiem użytkownicy Internetu współtworzą go i nadają mu obecny kształt¹¹. Dlatego też powstaje wiele inicjatyw, których celem jest ochrona Sieci przed pełnym skomercjalizowaniem jej funkcjonowania, i tworzenie kanałów swobodnej dystrybucji zasobów informacyjnych, szczególnie naukowych i dydaktycznych, które mają stanowić podstawę e-learningu i edukacji obywateli społeczeństwa wiedzy.

⁹ Google Search Book, pierwotnie Google Print, dostępny w World Wide Web: <http://books.google.com/>

¹⁰ Open Content Alliance, dostępny w World Wide Web: <http://www.opencontentalliance.org/>

¹¹ M. Castells, *Galaktyka Internetu. Refleksje nad Internetem, biznesem i społeczeństwem*, Poznań 2003, s. 40.

Najstarszą z nich jest ruch Open Access, który propaguje i tworzy wolny obieg informacji cyfrowej. Punktem wyjścia dla jego zwolenników jest fundamentalne założenie, że działalność naukowa i upowszechnianie jej wyników nie może być przedsięwzięciem nastawionym na zysk, ale przede wszystkim ma być służbą na rzecz ludzkości i jej rozwoju¹². Zmierzają oni do budowania alternatywnego, całkowicie wolnego i niezależnego obiegu informacji naukowej w Internecie. Z inicjatywy OA powstają więc bezpłatnie dostępne czasopisma naukowe, archiwa i repozytoria zawierające wiele materiałów, w tym również postprinty, które ukazały się już w komercyjnym obiegu wydawniczym¹³. Największym obecnie katalogiem indeksującym otwarte czasopisma jest Directory of Open Access Journals, który udostępnia już ponad 2500 tytułów z kilkudziesięciu dyscyplin naukowych, m.in. z informatyki, socjologii, ekonomii, prawa czy historii literatury¹⁴.

Szerokie działania na rzecz wolnego dostępu do cyfrowych dóbr kultury podejmuje również pozarządowa organizacja Creative Commons. Powołali ją do życia naukowcy i intelektualiści, w tym głównie prawnicy, którzy postrzegają Internet jako medium stwarzające niepowtarzalną szansę na grupowanie ludzi wokół idei tworzenia określonych dóbr, zasobów wiedzy, czy podejmowania konkretnych inicjatyw. Efekty ich pracy są natychmiast udostępniane online wszystkim zainteresowanym osobom, które korzystając z nich poszerzają grono wspólnoty. Dobra te są więc nieustannie wykorzystywane, modyfikowane i wytwarzane w wyniku pracy wielu ludzi, która nie ma charakteru wyłącznie komercyjnego. Twórcy CC nie dążą bynajmniej do likwidacji prawa, ale proponują jego rozsądną zmianę, by zachowana została równowaga pomiędzy ochroną własności intelektualnej, a dostępem do informacji. Zadanie to realizują poprzez propagowanie odpowiednich umów licencyjnych, uwzględniających różne warianty udostępniania określonego dzieła oraz eliminowanie pośredników pomiędzy jego twórcą i odbiorcą. W ten sposób stwarzają autorom dogodne warunki do udostępniania swoich dzieł w Internecie, do promowania swojej twórczości i pełnego wykorzystywania jej przez jak największą grupę użytkowników¹⁵.

¹² Zob. B. Niedźwiedzia, *Upowszechnianie wyników badań naukowych w Sieci*, „Forum Akademickie” 2005, nr 4, [on-line], [dostęp: 07.03.2007], dostępny w World Wide Web:

http://www.forumakad.pl/archiwum/2005/04/17-za-open_acces_cz_1.htm

¹³ Najstarsze i najbardziej znane archiwa reprintów powstały w USA i Europie, przykładem może być: CogPrints, udostępniający publikacje m.in. z zakresu psychologii, neurologii, biologii, matematyki, filozofii, dostępny w World Wide Web: <http://cogprints.org/>, arXiv, zawierający materiały z nauk ścisłych, dostępny w World Wide Web: <http://arxiv.org/>, czy RePEc, udostępniający publikacje z dziedziny ekonomii, dostępny w World Wide Web: <http://repec.org/>

¹⁴ Directory of Open Access Journals, dostępny w World Wide Web: <http://www.doaj.org/>

¹⁵ Creative Commons Polska, dostępny w World Wide Web: <http://pl.creativecommons.org/>, zob. także: L. Leasing, *Wolna kultura*, Warszawa 2005 lub [on-line], [dostęp: 07.03.2007], dostępny w World Wide Web: <http://www.futrega.org/wk/>

Polityka cyfryzacji

Istnieje już wiele krajowych i międzynarodowych organizacji, fundacji czy konsorcjów, w ramach których powstają różne programy i inicjatywy wspierające proces digitalizacji zbiorów bibliotecznych oraz tworzenia, rozwoju i upowszechniania bibliotek cyfrowych. Już w 1993 roku utworzony został przez UNESCO program „Pamięć Świata” (Memory of the World). Ma on na celu podejmowanie działań służących zachowaniu, ratowaniu i udostępnianiu dzieł najbardziej istotnych dla rozwoju naszej cywilizacji. Przedmiotem jego zainteresowania są dokumenty, rękopisy i druki o szczególnym znaczeniu historycznym, które są odpowiednio skanowane i przechowywane na nośnikach cyfrowych (CD, DVD)¹⁶.

Natomiast jednym z ostatnich projektów jest budowa Europejskiej Biblioteki Cyfrowej, zainicjowana przez Komisję Europejską w ramach programu i2010: biblioteki cyfrowe. Ma on objąć wszystkie biblioteki narodowe krajów europejskich, i sprawić, by bogate dziedzictwo kulturowe Europy można było łatwiej i w bardziej interesujący sposób wykorzystywać w Internecie. Projekt obejmuje trzy aspekty:

- digitalizację zbiorów analogowych,
- dostęp online, który jest warunkiem wstępnym dla maksymalizacji korzyści, jakie informacje mogą przynieść obywatelom, naukowcom i przedsiębiorstwom,
- konserwację i przechowywanie, które zapewnią przyszłym pokoleniom dostęp do materiałów cyfrowych oraz zapobiegną utracie najcenniejszych zasobów. Komisja Europejska, powołując specjalną Grupę Ekspertów, sprecyzowała praktyczne założenia dla budowy EBC, której zasób do końca 2010 r. ma liczyć przynajmniej sześć milionów zdigitalizowanych obiektów, pochodzących z europejskich bibliotek, archiwów i muzeów¹⁷.

Także w Polsce rozpoczął się proces digitalizacji zbiorów i tworzenia cyfrowych kolekcji. Coraz więcej bibliotek tworzy w swych strukturach pracownie digitalizacji bądź zleca cyfryzację części swoich zbiorów zewnętrznym firmom. Elektroniczne kopie udostępniane są online, tworząc coraz bogatszy polskojęzyczny zasób pełnotekstowy. Oprócz stworzonej w Uniwersytecie Gdańskim i wspieranej przez program UNESCO Biblioteki

¹⁶ Memory of the World, dostępny w World Wide Web: www.unesco.org/webworld, zob. także: E. Głowacka, *"Pamięć świata" - program UNESCO poświęcony ochronie i udostępnianiu najcenniejszych zasobów bibliotek i archiwów*, „Bibliotekarz” 2001, nr 4, s. 20.

¹⁷ *Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiego Komitetu Ekonomiczno-Społecznego oraz Komitetu Regionów i2010: Biblioteki Cyfrowe*, COM (2005) 465 wersja ostateczna [on-line]. Bruksela: Komisja Wspólnot Europejskich, 30 września 2005 [dostęp: 07.03.2007], dostępny w World Wide Web: http://ec.europa.eu/information_society/activities/digital_libraries/doc/communication/pl_comm_digital_libraries.pdf

Literatury Polskiej w Internecie¹⁸, zainicjowanej przez Bibliotekę Uniwersytecką w Warszawie Cyfrowej Kolekcji Czasopism Polskich¹⁹, czy rządowej Polskiej Bibliotece Internetowej²⁰, powstaje coraz więcej kolekcji regionalnych, jak np. Wielkopolska Biblioteka Cyfrowa²¹, Dolnośląska Biblioteka Cyfrowa²², czy Kujawsko-Pomorska Biblioteka Cyfrowa²³. By proces cyfryzacji w Polsce przebiegał spójnie i według jednolitych zasad, Minister Kultury i Dziedzictwa Narodowego powołał w kwietniu 2006 r. Zespół ds. digitalizacji, który ma „koordynować proces digitalizacji oraz udostępnienia zasobów dziedzictwa kulturowego i dorobku naukowego za pośrednictwem Internetu”²⁴.

Pracownia Digitalizacji w BG UKW

Rozpoczęcie procesu digitalizacji w Bibliotece Głównej Uniwersytetu Kazimierza Wielkiego w Bydgoszczy stało się możliwe dzięki pozyskaniu unijnych dotacji na budowę regionalnej biblioteki cyfrowej. Pracownia Digitalizacji uruchomiona została w czerwcu 2005 r. W związku z tym, że biblioteka nie posiadała wcześniej takiej jednostki w swojej strukturze, musiała na ten cel wygospodarować odpowiednio przystosowane do pracy przy komputerach i skanerze pomieszczenie. Wymagało ono m.in. zainstalowania lokalnej sieci komputerowej, która umożliwia przesyłanie danych zarówno pomiędzy jednostkami znajdującymi się w Pracowni, jak i kontakt z serwerem, gdzie archiwizowane są zeskanowane pliki. Konieczne było również odpowiednie zaciemnienie pomieszczenia, czego wymaga praca przy skanerze. Obecnie wyposażenie pracowni stanowi:

- skaner Zeutschel z oprogramowaniem Omnicam 5000 TT,
- płaski skaner Epson GT-15000,
- aparat cyfrowy Canon EOS 20D,
- aparat cyfrowy Canon A 520,
- statyw do aparatów Kaiser RTX,
- drukarka kolorowa Epson AcuLaser C1900,

¹⁸ Biblioteka Literatury Polskiej w Internecie, dostępna w World Wide Web: <http://univ.gda.pl/~literat/>

¹⁹ Cyfrowa Kolekcja Czasopism Polskich, dostępna w World Wide Web: <http://www.buw.uw.edu.pl/zasoby/ckcp.htm>

²⁰ Polska Biblioteka Internetowa, dostępna w World Wide Web: <http://www.pbi.edu.pl/index.html>

²¹ Wielkopolska Biblioteka Cyfrowa, dostępna w World Wide Web: <http://www.wbc.poznan.pl/>

²² Dolnośląska Biblioteka Cyfrowa, dostępna w World Wide Web: <http://www.dbc.wroc.pl/>

²³ Kujawsko-Pomorska Biblioteka Cyfrowa, dostępna w World Wide Web: <http://kpbc.umk.pl/>

²⁴ *Decyzja nr 12 Ministra Kultury i Dziedzictwa Narodowego z dnia 24 kwietnia 2006 r. w sprawie powołania Zespołu do spraw digitalizacji*, [on-line], [dostęp: 07.03.2007], dostępny w World Wide Web: <http://bip.mkidn.gov.pl/bip/document/?docId=576>

- trzy komputery z: oprogramowaniem Windows XP Professional, procesorem Intel Pentium 4 HT 3.0 GHz, ram 3 GB DDR i dyskiem 160 GB HDD.

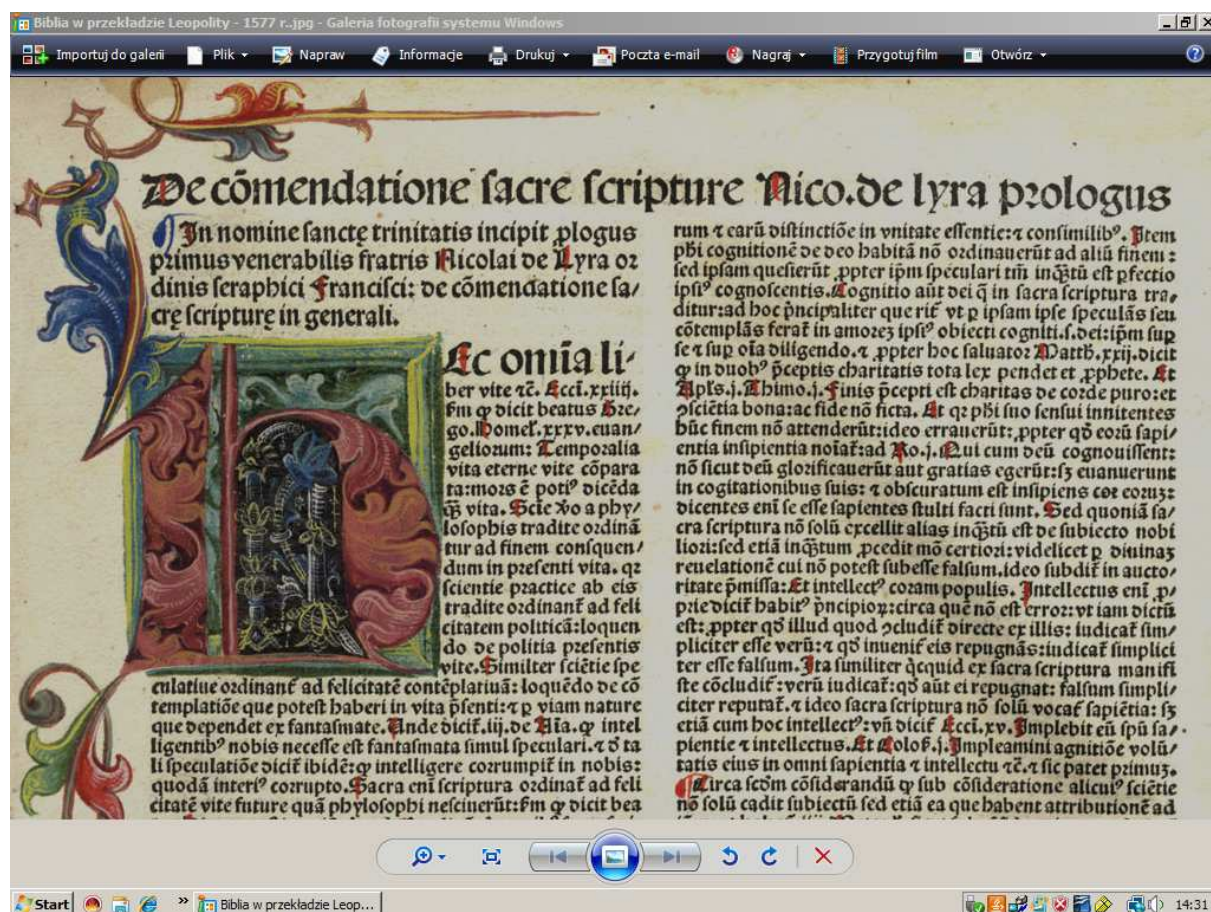
Digitalizacja zbiorów jest zadaniem stosunkowo nowym w bibliotekach, nie ma więc jeszcze dobrze określonego modelu organizacyjnego. Sprawność i efektywność całego procesu w dużym stopniu zależy więc od poprawnego zaplanowania wszystkich jego etapów. W chwili obecnej w BG UKW najważniejsze z nich, to:

- selekcja oraz typowanie dokumentów do skanowania,
- tworzenie list publikacji wyselekcjonowanych do cyfryzacji, zawierających m.in. informacje o zakresie digitalizacji, archiwizacji, publikacji w Internecie,
- dostarczenie publikacji do Działu Opracowania, gdzie przygotowywane są metadane,
- skanowanie,
- przetwarzanie plików w programach graficznych,
- konwersja dokumentów przeznaczonych do publikacji do formatu DjVu,
- przetwarzanie bitmapy na edytowalny tekst poprzez funkcję OCR,
- tworzenie publikacji i wykonywanie prac redakcyjnych w systemie dLibra,
- tworzenie listy dokumentów opublikowanych online dla Działu Opracowania, gdzie informacje o dostępie elektronicznym wprowadzane są do katalogu biblioteki,
- archiwizacja zeskanowanych i obrobionych plików na płytach DVD bądź dyskach twardych,
- opracowywanie statystyk,
- współpraca z regionalnymi bibliotekami w ramach realizacji Kujawsko-Pomorskiej Biblioteki Cyfrowej.

Selekcja zbiorów

Oczywiście, proces digitalizacji w bydgoskiej bibliotece nie obejmuje wszystkich materiałów w niej zgromadzonych. Nie zmierza on do stworzenia elektronicznej kopii całego zasobu i wyeliminowania jej dotychczasowej, tradycyjnej funkcji, ale raczej do uzupełniania i wzbogacania jej oferty. Jego celem jest ochrona zbiorów szczególnie cennych i równocześnie swobodne ich udostępnianie w wersjach cyfrowych, a przede wszystkim tworzenie kolekcji cyfrowych w Internecie. Ze zbiorów specjalnych biblioteki w pierwszym etapie wyselekcjonowanych zostało kilka starodruków przeznaczonych do pełnej digitalizacji, których cyfrowe wtórники przechowywane są na dyskach DVD. Są to m.in.: Jan Łaski, *Commune Incliti Poloniae Regni Privilegium*, Kraków 1506, Arystoteles, *Trium librorum de anima*, Kraków 1513, Maciej Miechowita, *Chronica Polonorum*, Kraków 1512, Stanisław

Sokołowski, *Justus Joseph*, Kraków 1586, Piotr Skarga *Kazania o siedmi sakramentach Kościoła S. Katolickiego*, Kraków 1600, Jan Kochanowski, *Fragmenta*, Kraków 1604, Jakub Sobieski, *Commentariorum Chotinensis Belli Libri Tres*, Gdańsk 1646. Zeskanowane zostały również wybrane miedzioryty Erika Dahlberga, pochodzące z dzieła Samuela Pufendorfa *De rebus a Carolo Gustavo Sueciae Rege gestis*, wydane w Norymberdze w 1696 r., które udostępnione zostały użytkownikom w formie wystawy w bibliotece cyfrowej.



Fragment Biblii z 1498 r. udostępnianej w BG UKW na DVD.

Znacznie trudniejsze jest określenie odpowiedniego kryterium doboru publikacji przeznaczonych do udostępniania online, by nie powstał zbiór przypadkowych dokumentów, które w niewielkim tylko stopniu będą wykorzystywane przez użytkowników²⁵. Na tym etapie wyraźnie brakuje koordynacji na poziomie ogólnopolskim oraz centralnego źródła

²⁵ Coraz bardziej żywa staje się koncepcja angażowania użytkowników bibliotek cyfrowych w proces ich tworzenia, m.in. poprzez współredagowanie metadanych czy właśnie dobór publikacji. Powstało już nawet Forum społeczności czytelników i bibliotekarzy cyfrowych Biblioteka 2.0, którego twórcy wpisują rozwój bibliotek cyfrowych w nurt filozofii Web 2.0, [on-line], [dostęp: 07.03.2007], dostępne w World Wide Web: <http://forum.biblioteka20.pl/>

informacji o obiektach zdigitalizowanych i przygotowywanych do takiego procesu w innych ośrodkach. Zakres formalny prowadzonej selekcji określają oczywiście obowiązujące przepisy prawne, natomiast zakres rzeczowy wyznaczają cztery główne kolekcje KPBC. W ich obrębie dokonywany jest dobór materiału na podstawie jego dydaktycznej i naukowej wartości oraz częstotliwości wykorzystywania. W proces selekcji prócz bibliotekarzy włączani się również pracownicy Uniwersytetu Kazimierza Wielkiego, którzy albo sami przygotowują listy publikacji do zeskanowania, albo z wcześniej przygotowywanych wybierają tytuły najbardziej wartościowe. W ten sposób udostępnione w sieci WWW zostały m.in.: *O duszy nauczycielstwa* Jana Władysława Dawida, *Pedagogika polska w zarysie* Władysława Seredyńskiego, *Podręczna encyklopedia pedagogiczna* Feliksa Kierskiego, *Dzieje wychowania i szkół w Polsce* Antoniego Karbowiaka, *Dokumenty kujawskie i mazowieckie przeważnie z XIII wieku* Bolesława Ulanowskiego, *Monumenta Poloniae Paleographica* Stanisława Krzyżanowskiego czy *Monumenta Poloniae Historica*.

W Pracowni Digitalizacji konwertowane na postać cyfrową są głównie materiały pochodzące ze zbiorów własnych biblioteki, ale nie tylko. Część publikacji pochodzi ze zbiorów Biblioteki Miejskiej im. Jana Kasprowicza w Inowrocławiu (*Dziennik Kujawski*), z Biblioteki Wyższego Seminarium Duchownego we Włocławku (*Monumenta Historica Dioeceseos Wladislaviensis*) oraz z Wojewódzkiej i Miejskiej Biblioteki Publicznej im. dr Witolda Bełzy w Bydgoszczy, z którą udało się nawiązać bardzo owocną współpracę. Dzięki niej konwertowane na postać cyfrową i udostępniane online są niezwykle interesujące materiały pochodzące ze zbiorów regionalnych bydgoskiej księżnicy. Publikacje te, ze względu na ich wartość i stan wymagający szczególnej ochrony, są bardzo rzadko udostępniane czytelnikom w oryginale. Dzięki kopii cyfrowej zostały przywrócone do powszechnego obiegu, i nadają bibliotece cyfrowej charakter niepowtarzalnej i specyficznej kolekcji. Obecnie w KPBC dostępnych jest już ponad 4500 publikacji pochodzących ze zbiorów WiMBP w Bydgoszczy. Są to m.in. książki: *Ilustrowany przewodnik po Bydgoszczy* Stanisława Łabędzińskiego, *Ilustrowany przewodnik po Inowrocławiu i Kujawach* Karola Kopca, *Bydgoszcz: kilka historycznych wspomnień o zamku i mieście* Ignacego Polkowskiego, *Powiat Nakielski w XVI stuleciu* Edmunda Celliera, *Album Grudziądza* z cyklem widoków rotograviurowych, *Przewodnik ilustrowany po Włocławku* Stanisława Jankowskiego, *Księgi Adresowe Miasta Bydgoszczy*, mapy i plany miasta Bydgoszczy i okolic: *Croquis der Umgegend von Bromberg* z 1854 roku, czy *Strassenplan d. Stadt Bromberg mit Strassenverzeichniss* z 1900 roku, oraz liczna kolekcja regionalnych czasopism: *Gazeta*

Digitalizacja i obróbka plików

Serce Pracowni Digitalizacji stanowią skanery bądź aparaty cyfrowe, które przetwarzają utrwalone na fizycznych nośnikach obrazy analogowe na postać numeryczną. W BG UKW główna praca wykonywana jest na skanerze książkowym firmy Zeutschel z oprogramowaniem przetwarzającym obraz Omniscan 5000 TT. Umożliwia on wykonywanie kopii cyfrowych z oryginału maksymalnie w formacie A2 (645 x 435 mm) w rozdzielczości 600 dpi, trybie 8-bitowym, w 256 odcieniach szarości. Wszystkie skany wykonywane są z reguły w odcieniach szarości, do formatu TIFF w rozdzielczości 300 dpi. Mimo, iż zajmują one nieco więcej pamięci niż pliki 1-bitowe, jednak jakość prezentacji zapisanej w większej głębi jest znacznie wyższa, a niekiedy w ogóle umożliwia odczytanie tekstu, który w przypadku plików 1-bitowych jest miejscami nieczytelny. Skaner wyposażony jest w zimne, pozbawione promieniowania UV, niemigotające oświetlenie, stół uchylony z automatyczną regulacją szklanej płyty dociskowej i „kołyskę” dostosowującą się do kształtu książki i dociskającą ją od spodu. Dołączone do skanera oprogramowanie uzupełnia cały proces cyfryzacji poprzez automatyczne numerowanie plików, zmianę orientacji zeskanowanych stron, czy wstępną obróbkę obrazów (przycinanie, maskowanie, usuwanie skośności, wzmacnianie kontrastu, podwyższanie ostrości, itd.).

Do digitalizacji dokumentów kolorowych wykorzystywany jest aparat cyfrowy Canon EOS 20D. Dzięki wyposażeniu w 8,2-megapikselową matrycę CMOS, precyzyjną automatykę ostrości i bogate menu umożliwia on uzyskanie bardzo dobrych efektów przy fotografowaniu cennych dzieł, a jego światło jest dla nich zupełnie bezpieczne. W ten sposób sfotografowane zostały m.in. iluminowane fragmenty pierwszego tomu *Biblia cum glosa ordinaria Walafridi Strabonis* z oficyny drukarskiej Jana Frobenia i Jana Petriego, wydane w Bazylei w 1498 roku, kilka niezwykle interesujących regionalnych map, np. *Plan der Stadt Bromberg* z 1876 roku, czy przedstawiająca kanał bydgoski z fragmentami Noteci *Canal vermittelt welchen die Netze und Brah vereinigt worden* z 1774 roku oraz wiele grafik, zdjęć i okładek książek. Do wykonywania zdjęć wykorzystywany jest także statyw, który nie tylko umożliwia utrzymanie aparatu w stabilnej pozycji, ale również daje możliwość uniesienia go na optymalną wysokość, bądź ustawiania pod odpowiednim kątem. Do digitalizacji barwnych dokumentów stosowany jest także płaski skaner biurowy Epson GT-

15000, jednak ze względu na szkodliwy wpływ jego światła wykorzystywany jest wyłącznie do dokumentów nowych i niewymagających szczególnej ochrony.



Obsługa skanera Zeuschel w Pracowni Digitalizacji BG UKW.

Otrzymane w wyniku skanowania „surowe” pliki w formacie TIFF są następnie poddawane obróbce graficznej. Przede wszystkim dla uzyskania lepszych walorów estetycznych, ale także z przyczyn ekonomicznych. Podstawowe zabiegi, jak kadrowanie, obracanie do odpowiedniego poziomu, czyszczenie oraz kontrastowanie, wykonywane są za pomocą programu Photoshop lub IrfanView. Czynności te są niezwykle czasochłonne, a ich sprawne wykonywanie w dużym stopniu zależy od technicznych możliwości jednostki, na której są wykonywane. Tak przygotowane pliki są następnie archiwizowane na płytach DVD i na dyskach twardych. Na dyskach optycznych przechowywane są głównie dokumenty cenne, wymagające szczególnej ochrony, które procesowi digitalizacji powinny być poddane tylko raz. Na jednej płycie zapisany jest jeden dokument zwarty bądź jeden miesiąc danego czasopisma, nawet jeżeli w ten sposób nie zostanie wykorzystana cała pojemność nośnika. Taka organizacja umożliwia jednak sprawne wyszukiwanie publikacji w coraz większym

zbiorze płyt, udostępnianie ich offline w czytelnich biblioteki oraz w przyszłości migrację na inny nośnik²⁶. Natomiast dokumenty, które przeznaczone są wyłącznie do udostępniania online archiwizowane są na dyskach twardych.

Zasadniczą cechą wszystkich dokumentów cyfrowych jest format, w jakim zostały one zapisane. Ogół stosowanych obecnie formatów podzielić można na dwie podstawowe grupy: archiwalne i prezentacyjne²⁷. W BG UKW do archiwizacji wykorzystywany jest format TIFF, natomiast do udostępniania publikacji w bibliotece cyfrowej format DjVu. Obecnie jest on zdecydowanie najbardziej efektywny spośród powszechnie stosowanych formatów graficznych. Jego istotę stanowi oryginalna metoda segmentacji obrazu oraz symultanicznie wykorzystywane algorytmy kompresji danych. Każdy plik dzielony jest na odrębne warstwy, i każda z nich z osobna poddawana jest procesowi optymalizacji. W efekcie dokumenty w formacie DjVu zajmują bardzo mało miejsca, są nawet dziesięciokrotnie mniejsze od plików wyjściowych. Mimo tak efektywnej kompresji zachowują niezwykle wysoką jakość obrazu i bardzo dobrą czytelność tekstu, porównywalną wręcz z plikiem źródłowym, zeskanowanym z rozdzielczością 300 dpi. Kolejnym walorem formatu DjVu jest technika częściowego odczytu danych. Dzięki jej zastosowaniu dokument pojawia się w przeglądarce od razu, zanim jeszcze pobrane zostaną wszystkie dane i progresywnie przybiera swoją ostateczną postać.

Do tworzenia i modyfikowania plików DjVu w BG UKW wykorzystywany jest program Document Express w wersji Professional 5.0. Posiada on również wbudowany moduł OCR, który w dużym stopniu przetwarza mapę bitową na edytowalny tekst. Jest to rzecz niezwykle istotna w przypadku tworzenia cyfrowego zasobu. Możliwość pełnotekstowego przeszukiwania, swobodnej wewnętrznej nawigacji i sprawnej orientacji w strukturze danej publikacji jest praktycznie standardem, który z jednej strony wyznaczają dokumenty analogowe, a z drugiej dokumenty cyfrowe powstałe od razu w wersji elektronicznej.

Kujawsko-Pomorska Biblioteka Cyfrowa

Szybki rozwój i sukces udostępniania dokumentów elektronicznych poprzez biblioteki cyfrowe w Stanach Zjednoczonych i w Europie Zachodniej zachęcił również środowiska bibliotekarskie i naukowe w Polsce do podejmowania działań na rzecz budowy własnych

²⁶ E. Stachowiak-Musiał, *Digitalizacja a ochrona zbiorów - współczesne tendencje*, w: *Aktualne tendencje ochrony zbiorów bibliotecznych i archiwalnych*, Warszawa 2002, s. 68.

²⁷ A. Trembowiecki, *Digitalizacja zbiorów bibliotecznych. Teoria i praktyka*, Warszawa 2006, s. 54.

zasobów w Internecie. Także w regionie kujawsko-pomorskim idea tworzenia kolekcji cyfrowych stawała się coraz bardziej żywa. Głównym problemem uniemożliwiającym jej realizację był brak odpowiednich funduszy. By je pozyskać z inicjatywy Biblioteki Uniwersyteckiej w Toruniu, władz UMK oraz rektorów wszystkich szkół wyższych województwa powstało w 2003 r. Konsorcjum Bibliotek Naukowych Regionu Kujawsko-Pomorskiego. Jednym z jego podstawowych zadań było wspólne stworzenie elektronicznego zasobu dokumentów publikowanych w Internecie²⁸. Opracowany został projekt biblioteki cyfrowej, stwarzającej wszystkim mieszkańcom regionu szybki dostęp do zasobów wiedzy oraz promującej historię i kulturę regionu w całym świecie poprzez sieć WWW. Inicjatywa zyskała poparcie władz województwa, dzięki czemu została włączona w projekt współfinansowany przez Unię Europejską z funduszy Zintegrowanego Programu Operacyjnego Rozwoju Regionalnego 2004-2006 oraz Regionalnego Programu Operacyjnego 2004-2006 województwa kujawsko-pomorskiego: Infrastruktura społeczeństwa informacyjnego, i otrzymała dotacje umożliwiające rozpoczęcie pracy²⁹.

Oficjalnie Kujawsko-Pomorska Biblioteka Cyfrowa została otwarta 2 grudnia 2005 r. Głównym jej koordynatorem jest Biblioteka Uniwersytecka w Toruniu, a współtworzy ją także Biblioteka Medyczna Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy Uniwersytetu Mikołaja Kopernika w Toruniu oraz Biblioteka Główna Uniwersytetu Kazimierza Wielkiego w Bydgoszczy. Efekt prac jest wspólny, użytkownik korzysta z jednej platformy, z jednej biblioteki, jednak cała praca odbywa się lokalnie w każdej z poszczególnych bibliotek. Dzięki temu każda z nich zachowuje dużą swobodę i autonomię działania. Każda ksiąznica posiada własną pracownię digitalizacji, samodzielnie skanuje wszystkie dokumenty, przetwarza otrzymane pliki, przygotowuje je do publikacji, kataloguje, przypisuje do odpowiedniej kolekcji, i wreszcie umieszcza gotowe publikacje w bibliotece cyfrowej określając prawa dostępu do każdej z nich. Także selekcja zbiorów przeznaczonych do skanowania odbywa się lokalnie, w każdej z bibliotek z osoba. Oczywiście, niezbędne było ustalenie wspólnych reguł i zasad obowiązujących wszystkie biblioteki partycypujące w

²⁸ M. Kempa, *Konsorcjum Bibliotek Naukowych Regionu Kujawsko-Pomorskiego - powstanie, zarys działalności, projekty na przyszłość*, „Biuletyn EBIB” 2004, nr 10, [on-line], [dostęp: 07.03.2007], dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2004/61/kempa.php>.

²⁹ D. Czyżak, *Kujawsko-Pomorska Biblioteka Cyfrowa – stan zaawansowania realizacji projektu ZPORR*, „Biuletyn EBIB” 2005, nr 9, [on-line], [dostęp: 07.03.2007], dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2005/70/czyzak.php>, zob. także D. Czyżak, G. Szturo, M. Rakowska, *Kujawsko-Pomorska Biblioteka Cyfrowa - problemy i zagadnienia związane z realizacją I etapu jej rozwoju*, „Biuletyn EBIB” 2006, nr 11, [on-line], [dostęp: 07.03.2007], dostępny w World Wide Web: http://www.ebib.info/2006/81/a.php?czyzak_szturo_rakowska

projekcie. Przede wszystkim dotyczyły one jednolitego sposobu opisywania publikowanych dokumentów i tworzenia zestawu metadanych.

dLibra

Do zarządzania KPBC wybrany został system dLibra, stworzony i niezwykle dynamicznie rozwijany przez Poznańskie Centrum Superkomputerowo-Sieciowe³⁰. Jest to program, który obejmuje wszystkie etapy pracy: zarządzanie, przechowywanie, przeszukiwanie i udostępnianie publikacji. Jego głównymi komponentami są aplikacje obsługujące trzy kategorie użytkowników: administratorów, redaktorów i czytelników. Administratorzy i redaktorzy systemu, w tym również katalogerzy, korzystają z własnych aplikacji umożliwiających im pełne zarządzanie zasobem. Wszystkie dokumenty w niej przechowywane ułożone są w hierarchiczną strukturę katalogów. Każda biblioteka posiada swój własny katalog, który może zawierać dowolną liczbę podkatalogów i publikacji wprowadzanych przez jej redaktora, i może być dowolnie przez niego modyfikowany. Jest to struktura wewnętrzna biblioteki służąca do sprawnego zarządzania całą kolekcją, nie związana ze sposobem udostępniania publikacji użytkownikom.

Metadane

Tworzenie publikacji w systemie dLibra odbywa się etapami. Najpierw przez katalogerów wprowadzane są metadane. W ten sposób powstaje publikacja planowana. Na tym poziomie pracy użytkownik biblioteki widzi już szczegółowe dane bibliograficzne obiektów przygotowywanych do publikacji, dzięki czemu wie, jakie tytuły będą niebawem w bibliotece dostępne. W BG UKW redagowaniem opisów bibliograficznych zajmują się cztery osoby: dwie opracowują publikacje zwarte i wielotomowe, jedna zbiory specjalne i jedna czasopisma. Do opisu publikacji w systemie dLibra wykorzystywany jest format Dublin Core w wersji 1.1., który stanowi obecnie standard dla obiektów publikowanych w sieci WWW. Składa się on z zestawu 15 prostych atrybutów, które można łatwo konfigurować i w zależności od potrzeby dostosować do wymagań konkretnej biblioteki. Ilość atrybutów wykorzystywanych do opisu publikacji może być dowolna, a każdy z nich może zawierać

³⁰ Opis funkcjonalności systemu, informacje techniczne, dokumentacja, wersja testowa, forum użytkowników i administratorów systemu i wiele innych informacji dostępnych jest na stronie głównej projektu dLibra, dostępnej w World Wide Web: <http://dlibra.psnc.pl/>

wiele wartości³¹. Niestety, w chwili obecnej nie ma jeszcze ogólnopolskich przepisów katalogowania, które jednoznacznie precyzowałyby sposób tworzenia opisów bibliograficznych w kolekcjach cyfrowych. Na potrzeby KPBC ustalony został schemat metadanych rozszerzony do 20 elementów, który obowiązuje wszystkie biblioteki współtworzące jej zasób. Dublin Core zawiera wiele elementów, które w ogóle nie występują w standardowym opisie bibliograficznym tworzonemu na potrzeby katalogu bibliotecznego, jak np.: format prezentacji danych, miejsce digitalizacji, czy lokalizacja przechowywania oryginału. Część z nich generowana jest automatycznie przez system na podstawie zawartych w nim informacji.

Każdy opis sporządzany jest na podstawie oryginalnego dokumentu. Jeżeli tylko jest to możliwe, to jest on konwertowany z formatu MARC, z katalogu biblioteki UKW, tworzonego w systemie Horizon. Dotyczy to wszystkim publikacji zwartych oraz tych dokumentów, które są opracowywane w katalogu elektronicznym. Jeżeli określona pozycja nie jest jeszcze reprezentowana w Horizonie, to opis taki sporządzany jest najpierw w katalogu bibliotecznym, a następnie konwertowany do dLibry. Taka procedura pozwala zachowywać optymalną poprawność sporządzanych opisów i dążyć do ujednolicania formy stosowanych atrybutów. Dzięki temu pośrednio wykorzystywana jest Kartoteka Haseł Wzorcowych NUKAT, z którym biblioteka UKW aktywnie współpracuje. W systemie dLibra wszystkie wartości każdego z poszczególnych atrybutów są grupowane i przechowywane w słowniku, do którego osoby tworzące metadane mają w każdej chwili dostęp. W efekcie raz wprowadzone do bazy atrybuty mogą być potem wielokrotnie wykorzystywane, co znacznie ułatwia pracę i umożliwia stosowanie ujednoliconych form. Słownik wartości atrybutów umożliwia również stosowanie synonimów słów kluczowych. Ma to ogromne znaczenie dla użytkowników biblioteki, którzy poszukując określonych materiałów w jej coraz bogatszych zasobach mogą w sposób prosty i jednoznaczny dotrzeć do wszystkich interesujących ich publikacji, które są w bibliotece udostępniane. Jest to również niezwykle istotne w przypadku wyszukiwania rozproszonego w sieci bibliotek cyfrowych, których obecnie powstaje coraz więcej.

Opisy bibliograficzne dokumentów, które nie są jeszcze opracowywane w formacie MARC, jak np. starodruki czy grafiki, są tworzone bezpośrednio w systemie dLibra. W przypadku publikacji grupowych, czyli wydawnictw ciągłych bądź wielotomowych, konwertowany jest tylko opis główny, zbiorczy, natomiast rekordy poszczególnych tomów

³¹ M. Nahotko, *Metadane: sposób na uporządkowanie Internetu*, Kraków 2004, s. 122.

czy numerów dziedziczą niezmiennie elementy opisu głównego. Wprowadzane są tylko dane właściwe dla określonego woluminu, jak data wydania, rocznik, numer, itd. Użytkownik widzi od razu całą strukturę publikacji i może swobodnie przejść do wybranego tomu czy numeru, a następnie wyedytować jego pełen opis bibliograficzny.

Do każdego zestawu metadanych przechowywanego w systemie dLibra można również wprowadzić informacje dodatkowe, komentarz czy uwagi, które w większym stopniu umożliwiają jednoznaczną identyfikację określonego dokumentu, a nie są częścią opisu bibliograficznego. Standardowo każda publikacja umieszczona w KPBC posiada także dokładne informacje o tym, kiedy została w bibliotece umieszczona, kiedy miała miejsce jej ostatnia modyfikacja oraz ile razy, od czasu publikacji, jej treść była wyświetlana. Na bazie tych informacji tworzone są statystyki, dostępne również użytkownikom biblioteki na jej stronie głównej. Ponadto system dLibra umożliwia wprowadzanie informacji wewnętrznych, wykorzystywanych przez administratorów czy redaktorów biblioteki, a niewidocznych dla jej użytkowników.



Gazeta Bydgoska dostępna w zasobie KPBC.

Kolejnym etapem pracy przy tworzeniu publikacji jest nadawanie jej określonych praw odczytu. Każdemu obiektowi znajdującemu się w bibliotece musi być przypisany określony ich rodzaj, by w ogóle został on udostępniony czytelnikom. Uzależnione jest to głównie od sytuacji prawnej danej publikacji. Kategoria Public Group obejmuje materiały wolne od praw autorskich i wydawniczych, i otwiera dostęp do nich wszystkim użytkownikom biblioteki. W przypadku publikacji prawnie chronionych autor, wydawca, czy obecny właściciel praw autorskich może określić obszar dostępu do danych tytułów, ograniczając go wyłącznie do określonej grupy użytkowników. Przykładem może być *Bibliografia Polska* Karola Estreichera, która jest dostępna w KPBC tylko dla użytkowników sieci uczelnianej UMK. Istnieje również możliwość zdefiniowania w systemie użytkownika, który nie tylko otrzyma prawo do przeglądania określonej publikacji, ale także do zarządzania nią, czyli np. tworzenia nowego, zmodyfikowanego wydania bądź przyznawania prawa dostępu innym użytkownikom.

Wyszukiwanie publikacji

Każdy dokument wprowadzany do KPBC przypisany jest do określonej kolekcji, jednej lub kilku równocześnie. Każda z głównych kolekcji dzieli się na mniejsze, podkolekcje, które w zależności od potrzeby mogą być dodawane bądź modyfikowane przez redaktora biblioteki, odpowiedzialnego za strukturę całego zasobu. I tak kolekcja Regionalia dzieli się na Cuiaviana i Pomerania oraz Vilniana, a te z kolei na: Artykuły, Czasopisma regionalne, Druki muzyczne, Fotografie, Grafikę, Książki, Mapy, Pocztówki, itd. Kolekcja obejmująca Dziedzictwo kulturowe dzieli się na podkolekcje: Czasopisma, Książki XIX - XX w., Mapy, Materiały ulotne, Rękopisy, Zbiory emigracyjne oraz Zbory ikonograficzne. Natomiast kolekcja zawierająca Materiały Dydaktyczne dzieli się na poszczególne dyscypliny naukowe: Archeologię, Filozofię, Historię, Sztukę, itd. Dzięki temu każdy użytkownik biblioteki może szybko zorientować się, jaka jest rzeczowa zawartość całego zbioru. Kiedy wybierze określoną kolekcję czy podkolekcję, otrzyma informacje o ilości przypisanych do niej publikacji. Wszystkie może wyedytować w postaci uszeregowanej alfabetycznie listy tytułów, i z niej przejść bezpośrednio do opisu bibliograficznego wybranej publikacji.

Oczywiście użytkownicy mogą także korzystać ze standardowej wyszukiwarki, wybierając określone kryterium, według którego chcą przeglądać zasób bazy. Dzięki temu, że większość dokumentów publikowanych w KPBC poddawana jest procesowi OCR, istnieje możliwość prowadzenia poszukiwań nie tylko w opisie bibliograficznym, ale także w treści

publikacji. W efekcie użytkownik otrzymuje listę wyników spełniających warunek, z której może wybrać określoną publikację. Możliwe jest także redagowanie rozbudowanej kwerendy korzystając z operatorów logicznych Boola, bądź korzystanie z dostępnych na stronie głównej indeksów. Szeregują one zasoby biblioteki alfabetycznie według trzech atrybutów: autora, tytułu i słów kluczowych. Użytkownicy biblioteki otrzymują więc wiele możliwości dotarcia do opublikowanych w niej materiałów.

Ogromną zaletą systemu dLibra jest także to, że metadane dokumentów publikowanych w bibliotece cyfrowej są w pełni otwarte i czytelne dla systemów indeksujących zasoby sieci globalnej. Wszystkie publikacje bowiem zawierają równocześnie opis w formacie RDF (Resource Description Framework), który jest specyfikacją metadanych dla obiektów dostępnych w Web, implementowanych z reguły w języku XML. Umożliwia to ich automatyczną wymianę z systemami zewnętrznymi oraz współpracę z wyszukiwarkami internetowymi. Dzięki temu każdy dokument umieszczony w kolekcji KPBC jest indeksowany np. przez Google. W efekcie potencjalny użytkownik może sprawnie wyszukać materiały opublikowane w bibliotece cyfrowej, w ogóle nie wiedząc wcześniej o jej istnieniu.

Jeszcze innym sposobem docierania do materiałów opublikowanych w bibliotece cyfrowej jest zamieszczanie ich adresów w innych źródłach informacji, tzn. w katalogach, bazach danych, przypisach, bibliografiach, itd. Każda publikacja umieszczona w zasobach KPBC otrzymuje swój własny, unikalny identyfikator, który pozostaje niezmienny przez cały czas jej istnienia. Adres URL publikacji konstruowany jest na podstawie tego właśnie identyfikatora, dzięki czemu zawsze będzie on prowadził do tego samego dokumentu, niezależnie od jego ewentualnych zmian, wprowadzania kolejnych wydań czy nawet zupełnego usunięcia z biblioteki. W takim przypadku użytkownik otrzyma informację mówiącą o tym, że dana pozycja była udostępniana w bibliotece, ale z określonych przyczyn została z niej usunięta. Mimo więc niezwykle dynamicznej natury Internetu gwarantuje to dużą stabilizację, i umożliwia bezpieczne umieszczanie adresu w innych miejscach. W Bibliotece UKW informacje o udostępnieniu danego tytułu online są wprowadzane do rekordu bibliograficznego tej pozycji w katalogu bibliotecznym. W polu uwag dodawana jest informacja o wersji cyfrowej dokumentu, a w polu 856 adres publikacji. Dzięki temu użytkownicy katalogu, korzystający z niego poprzez Internet, mogą przejść od razu z poziomu opisu bibliograficznego do publikacji dostępnej w KPBC. Wymaga to niestety dużego nakładu pracy, konieczne bowiem jest modyfikowanie każdego rekordu z osobna, jednak daje bardzo wymierne efekty. Użytkownicy katalogu bibliotecznego otrzymują pełną informację o dostępie do materiału źródłowego, który jest już opublikowany online.

100	1 _	\$a Dawid, Jan Władysław \$d (1859-1914).
245	1 0	\$a O duszy nauczycielstwa / \$c napisał J. Wł. Dawid.
260	_ _	\$a Warszawa : \$b Nasza Księgarnia, \$c 1927.
300	_ _	\$a 40 s. ; \$c 20 cm.
530	_ _	\$a Dostępny również w formie elektronicznej.
856	4 1	\$u http://kpbc.umk.pl/publication/498

Rekord MARC umożliwiający bezpośredni dostęp z opisu bibliograficznego w katalogu bibliotecznym do pełnego tekstu opublikowanego w KPBC.

Zasób rozproszony

Polskich bibliotek cyfrowych udostępniających swoje kolekcje online powstaje coraz więcej, a ich zasoby rosną z każdym dniem. Niewątpliwie istotną korzyścią jest fakt, że w zdecydowanej większości pracują one w tym samym systemie, poznańskiej dLibrze. Pierwsza biblioteka oparta na niej, WBC, uruchomiona została z inicjatywy poznańskiego środowiska akademickiego w 2002 r. Obecnie funkcjonuje już w Internecie 16 bibliotek cyfrowych zarządzanych przez dLibrę. Prócz WBC i KPBC są to m.in.: Śląska Biblioteka Cyfrowa³², Zielonogórska Biblioteka Cyfrowa³³, Biblioteka Cyfrowa Uniwersytetu Wrocławskiego³⁴ czy Polona, Cyfrowa Biblioteka Narodowa³⁵. W przypadku zbiorów cyfrowych tradycyjny paradygmat biblioteki, jako instytucji udostępniającej własne zbiory, ulega zmianie. Wszystkie zasoby dostępne online, szczególnie z punktu widzenia ich użytkowników, można traktować jako jedną bibliotekę cyfrową, której zasób jest gromadzony, opracowywany i udostępniany w sposób rozproszony³⁶.

Jednak by było to możliwe, konieczne jest stosowanie jednolitych i spójnych metod redagowania metadanych oraz sposobu przeglądania zasobów. Dla użytkowników bowiem najważniejszy jest prosty i jednoznaczny sposób dostępu do gromadzonych materiałów. Poznański system taką możliwość oferuje, bowiem od wersji 2.2 posiada zaimplementowany mechanizm wyszukiwania rozproszonego, oparty na protokole OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Protokół ten służy do selektywnego

³² Śląska Biblioteka Cyfrowa, dostępna w World Wide Web: <http://www.sbc.org.pl/dlibra>

³³ Zielonogórska Biblioteka Cyfrowa, dostępna w World Wide Web: <http://zbc.uz.zgora.pl/dlibra>

³⁴ Biblioteka Cyfrowa Uniwersytetu Wrocławskiego, dostępna w World Wide Web: <http://www.bibliotekacyfrowa.pl/dlibra>

³⁵ Cyfrowa Biblioteka Narodowa "Polona", dostępna w World Wide Web: <http://www.polona.pl/dlibra>

³⁶ M. Nahotko, dz. cyt., s. 38.

udostępniania metadanych zgromadzonych obiektów zewnętrznym systemom. Umożliwia ich automatyczną wymianę, uwarunkowaną jednak stosowaniem standardu Dublin Core. W efekcie użytkownik korzystając z jednej, lokalnej biblioteki cyfrowej otrzymuje możliwość przeszukiwania zasobu wszystkich bibliotek cyfrowych, tworzących wspólną sieć³⁷. Wystarczy, że wprowadzając zapytanie zaznaczy opcję „przeszukaj zdalne biblioteki”, a otrzyma pełną informację o wszystkich dostępnych publikacjach. Na początku 2007 r. w KPBC dostępnych jest ponad 13 000 publikacji, z czego 6 000 zdigitalizowanych w BG UKW. Natomiast łączna liczba wszystkich publikacji zgromadzonych w bibliotekach cyfrowych opartych na systemie dLibra przekroczyła już 50 000.

Zakończenie

Dzisiaj nikt już nie stawia pytania o sens digitalizacji zbiorów bibliotecznych, jest to rzecz oczywista, także dla pracowników BG UKW. Pytania dotyczą raczej tego, jak przeprowadzać i udoskonalać cały proces, w jaki sposób prezentować otrzymane pliki użytkownikom i gdzie pozyskiwać niezbędne do tego zadania fundusze. Na całym świecie powstają kolejne projekty digitalizacji i pracują już tysiące skanerów, powołując do cyfrowego życia kolejne publikacje. Wszystkie te działania łączy jeden cel - ochrona najcenniejszych obiektów kultury oraz udostępnianie ich treści w zasobach Internetu. Dzięki temu drukowany zapis kultury staje się dostępny użytkownikom epoki cyfrowej, a biblioteki, tworząc cyfrowe kolekcje, pozostają wierne swojemu powołaniu, stojąc na straży wolnego i demokratycznego dostępu do wiedzy.

³⁷ T. Kalota, *Marzenie o polskim systemie rozproszonych bibliotek cyfrowych*, „Biuletyn EBIB” 2006, nr 4, [online], [dostęp: 03.02.2007], dostępny w World Wide Web: <http://www.ebib.info/2006/74/kalota.php>.